

Research Article

Rolling Bearing Fault Diagnosis Using a Deep Convolutional Autoencoding Network and Improved Gustafson–Kessel Clustering

Yaochun Wu ^{1,2}, Rongzhen Zhao ¹, Wuyin Jin,¹ Linfeng Deng,¹ Tianjing He ¹, and Sencai Ma¹

¹School of Mechanical and Electrical Engineering, Lanzhou University of Technology, Lanzhou 730050, Gansu, China

²School of Mechanical Engineering, Anyang Institute of Technology, Anyang 455000, Henan, China

Correspondence should be addressed to Rongzhen Zhao; zhaorongzhen@lut.cn

Received 10 May 2020; Revised 17 September 2020; Accepted 21 September 2020; Published 19 October 2020

Academic Editor: José J. Rangel-Magdaleno

Copyright © 2020 Yaochun Wu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Deep learning (DL) has been successfully used in fault diagnosis. Training deep neural networks, such as convolutional neural networks (CNNs), require plenty of labeled samples. However, in mechanical fault diagnosis, labeled data are costly and time-consuming to collect. A novel method based on a deep convolutional autoencoding network (DCAEN) and adaptive nonparametric weighted-feature extraction Gustafson–Kessel (ANW-GK) clustering algorithm was developed for the fault diagnosis of bearings. First, the DCAEN that is pretrained layer by layer by unlabeled samples and fine-tuned by a few labeled samples is applied to learn representative features from the vibration signals. Then, the learned representative features are reduced by t-distributed stochastic neighbor embedding (t-SNE), and the low-dimensional main features are obtained. Finally, the low-dimensional features are input ANW-GK clustering for fault identification. Two datasets were used to validate the effectiveness of the proposed method. The experimental results show that the proposed method can effectively diagnose different fault types with only a few labeled samples.

1. Introduction

Rolling bearings are crucial components widely used in modern machines [1]. They often survive in harsh working environments where they are frequently damaged. A sudden failure of the rolling bearings may result in unexpected downtime, significant economic losses, and casualties [2]. Therefore, it is meaningful to develop intelligent fault diagnosis methods for rolling bearings. In general, intelligent fault diagnosis of the bearing comprises two parts: feature extraction and condition identification [3, 4]. Traditional feature extraction methods based on the EMD, entropy, and multifractal method have been successfully used in mechanical fault diagnosis. Chen et al. [5] used EMD to decompose the bearing vibration signals and calculated the permutation entropy (PE) of the first few IMFs as the characteristic vector, and SVM was applied for operation status identification. Li et al. [6] applied a multifractal method to extract the generalized dimensional spectral

features from the vibration signals of hydropower units, and probabilistic neural network was used for fault diagnosis. However, these methods are largely dependent on prior knowledge about signal processing techniques and expert diagnosis experience.

As a rising star in the field of intelligent fault diagnosis, deep learning has received much attention in recent years [7–9]. Deep learning [10–13] can learn representative features hidden in the original data and directly establish an accurate mapping relationship between a model and the operating state of devices. The stacked autoencoder (SAE), or stacked denoising auto-encoder (SDAE), and CNN are two typical deep learning models. SDAE [14] is constructed by stacking multilayer denoising autoencoders. The representative features are extracted on unlabeled samples in a layerwise learning method. The original signal is fed into the first autoencoder to generate a latent representation, and this “code” is used to reconstruct the input signal. The output features of the first autoencoder are input into the second

one to further extract hierarchical representative features. The high-level representations are generated in an unsupervised way, which avoids the dependence on knowledge and experience. Lu et al. [15] established an SDAE with a transmitting rule of greedy training to learn high-order feature representations from the input data, and a softmax regression algorithm was employed for multiclass classification. Chen and Li [16] proposed an SDAE to extract features from the frequency domain information of the vibration signal for fault detection. These unsupervised feature extraction methods based on an SDAE have achieved remarkable results in fault diagnosis. However, the hidden layers in an SDAE are fully connected, which makes it difficult to train a large number of parameters with limited samples.

In contrast, a CNN has the characteristics of a local connection, weight sharing, and pooling structure, which reduces the parameters of the model and improves the training efficiency. Therefore, it is widely used in fault diagnosis. Janssens et al. [17] utilized a CNN model for the intelligent fault diagnosis of bearings and compared the advantages of a CNN with manually engineered features. Zhang et al. [18] proposed a deep WDCNN model to learn the deep representative features from the original bearing signals and achieved a high bearing fault recognition rate under variable loads and strong background noise. Jing et al. [19] constructed a CNN to learn deep features from the spectrum of vibration signals collected from a gearbox and realized the fault diagnosis of the gearbox. Jiang et al. [20] designed a multiscale CNN architecture for extracting multiscale features from the vibration signals and realized the fault diagnosis of wind turbine gearboxes. These CNN-based models all exhibited an excellent performance for learning and fault recognition. However, one problem we must face is that plenty of labeled samples are required for training the model. Labeling samples, however, is an expensive and time-consuming activity.

Cluster analysis is an effective method to deal with the classification of unlabeled data [21]. GK clustering [22] is a clustering algorithm based on the objective function, which can identify the extracted features. Benyounes et al. [23] applied GK clustering to identify the control parameters in a gas turbine and developed a reliable nonlinear mathematical model. Hua et al. [24] introduced a method for the driver intention classification by GK clustering. Li et al. [25] used GK clustering as one of four main clustering algorithms and introduced how to select the most suitable identification method for bearing fault diagnosis. Chen et al. [26] extracted multiscale permutation entropy and adopted GK clustering to detect rolling bearing faults. However, the inadequacy of the GK clustering method is that the different contributions of the data for clusters are not considered, and the number of cluster centers must be given in advance.

To overcome the abovementioned problems, we propose a novel method using the DCAEN and ANW-GK clustering for the intelligent fault diagnosis of bearings. Our contributions can be summarized as follows:

- (1) A DCAEN that was pretrained by unlabeled samples and fine-tuned by a few labeled samples was

constructed to learn the representative features from the original vibrational signals.

- (2) An improved ANW-GK clustering was proposed. The contribution of each sample for clusters was redefined using nonparametric weighted-feature extraction (NWFE) [27]. The initial number of clusters of the ANW-GK was adaptively determined using the PBMF function [28].
- (3) A novel bearing intelligent fault diagnosis method using the DCAEN and ANW-GK clustering was developed.

The rest of the paper is organized as follows. Section 2 briefly introduces the CAE, GK clustering, and NWFE. Section 3 describes the constructed DCAEN, improved ANW-GK clustering, and general procedure of the proposed method. In Section 4, the effectiveness of the proposed method is validated on two bearing datasets: one is an open benchmark dataset and the other is a laboratory-measured dataset. Section 5 discusses the proposed method. Finally, our findings and outcomes are summarized and elucidated in Section 6.

2. Theoretical Background

Some basic theories of the CAE, GK clustering, and NWFE are briefly introduced in Sections 2.1, 2.2, and 2.3. In Section 2.4, three kinds of clustering evaluation indexes are introduced.

2.1. Convolutional Autoencoder (CAE). CAE [29, 30] is an unsupervised learning model where the convolutional structure is embedded into the basic encoder. When the convolutional structure is used to replace the fully connected structure of the basic encoder, the encoder has the characteristics of sharing the local receptive field and the weight. As depicted in Figure 1, the CAE consists of the encoder and decoder network. The encoder comprises a convolutional layer and a pooling layer, which can transform the input data from a high-dimensional space into a set of 1d feature maps. The decoder comprises an unpooling layer and a deconvolutional layer, which can reconstruct the input data from the 1d feature maps. The parameters of the encoder and decoder are optimized by minimizing the difference between the reconstructed data and input data. Therefore, the CAE is a data-driven unsupervised feature extraction model.

2.1.1. Encoder. Given a high-dimensional input data X , a set of 1d feature maps $Y = \{y_1, y_2, \dots, y_H\}$, H is the number of convolution kernels, and the encoding process can be expressed as follows:

$$y_k = \psi(\sigma(K^{(k)} * X + b_k)), \quad k = 1, 2, \dots, H, \quad (1)$$

where $*$ is the convolution operation; $K^{(k)}$ is the k th convolution kernel; b_k is the bias of the k th convolution kernel; $\sigma(\cdot)$ is the exponential linear unit function (ELU); and $\psi(\cdot)$ is the max pooling with step s .

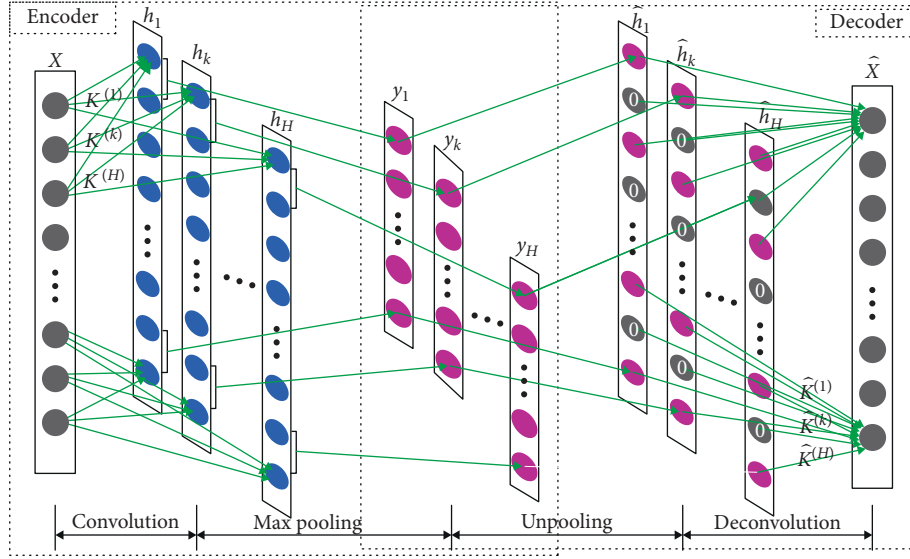


FIGURE 1: Structure of CAE.

2.1.2. *Decoder.* The decoder is used to reconstruct the input data X by the 1d feature maps y_k . The reconstructed data \hat{X} is expressed as follows:

$$\hat{X} = \sigma \left(\sum_{k \in H} \hat{K}^{(k)} * \hat{\psi}(y_k) + \hat{b}_k \right), \quad (2)$$

where $\hat{\psi}(\cdot)$ is the unpooling, which is used to expand the feature map according to the pooling step s of the encoder, that is, insert $s - 1$ zeros between the elements; $\hat{K}^{(k)}$ is the k th deconvolution kernel; and \hat{b}_k is the bias of the k th deconvolution kernel.

2.1.3. *Training.* The parameter set $\{K, b, \hat{K}, \hat{b}\}$ of the CAE is optimized by minimizing the reconstruction error. The error is defined as follows:

$$L(K, b, \hat{K}, \hat{b}) = \|\hat{X} - X\|^2. \quad (3)$$

2.2. *Gustafson-Kessel Clustering (GK Clustering).* The GK clustering algorithm [22] obtains the fuzzy membership matrix $U = [\mu_{ik}]_{c \times n}$ and cluster center $V = \{v_1, v_2, \dots, v_c\}$ by minimizing the objective function. Here, c is the number of clusters; n is the number of samples; and μ_{ik} is the fuzzy membership degree of the sample point k relative to the cluster center i . $\mu_{ik} \in [0, 1]$, $\sum_{i=1}^c \mu_{ik} = 1$, $1 \leq i \leq c$, and $1 \leq k \leq n$.

Given a clustering sample set $Z = \{Z_1, Z_2, \dots, Z_n\}$, its objective function is defined as

$$J(Z, U, V) = \sum_{i=1}^c \sum_{k=1}^n (u_{ik})^m D_{ik}^2, \quad (4)$$

$$D_{ik}^2 = (Z_k - v_i)^T \cdot A_i \cdot (Z_k - v_i), \quad (5)$$

$$A_i = \det(F_i)^{(1/m)} F_i^{-1}, \quad (6)$$

$$P_i = \left(\frac{1}{n}\right) \sum_{k=1}^n u_{ik}, \quad (7)$$

where m is the fuzzy index, generally, $m = 2$; D_{ik}^2 is the distance from any sample Z_k to the cluster center v_i , which is a square inner product norm; A_i is a positive definite symmetric matrix, determined by the clustering covariance matrix F_i ; and P_i is the prior probability of the i th cluster.

The Lagrange multiplication is used to optimize the objective function (equation (4)), and the necessary conditions for the minimum value of equation (4) are

$$u_{ik} = \frac{1}{\sum_{g=1}^c (D_{ik}/D_{kg})^{2/(m-1)}}, \quad (8)$$

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m Z_k}{\sum_{k=1}^n (u_{ik})^m}, \quad (9)$$

Then, the iteration algorithm of the GK clustering is described as follows:

- (1) The number of clusters c , the fuzzy index m , and the fuzzy membership matrix U are initialized. The iteration number is set as $l = 0, 1, \dots$
- (2) The cluster center v_i is calculated according to equation (9).
- (3) The covariance matrix F_i is calculated according to the following equation:

$$F_i = \frac{\sum_{k=1}^n (u_{ik})^m (Z_k - v_i)(Z_k - v_i)^T}{\sum_{k=1}^n (u_{ik})^m}. \quad (10)$$

- (4) The distance norm D_{ik}^2 is calculated according to equations (5) and (6).

(5) The fuzzy membership matrix U is updated according to equation (8).

For any positive η , if $\|U^{(l+1)} - U^{(l)}\| < \eta$, the operation is terminated. Otherwise, the number of iterations is increased, and $l \leftarrow l + 1$ is repeated until the condition is satisfied.

2.3. Nonparametric Weighted-Feature Extraction (NWFE). The NWFE [27] is used for computing different weighted clustering centers for each sample. The distributed weighting matrix is defined by the Euclidean distance between the samples and clustering centers. The nonparametric within-class scatter matrix S_w is defined as follows:

$$S_w = \sum_{i=1}^c P_i \sum_{k=1}^{n_i} \frac{w_k^{(i,j)}}{n_i} (Z_k^{(i)} - M_j(Z_k^{(i)})) \cdot (Z_k^{(i)} - M_j(Z_k^{(i)}))^T. \quad (11)$$

The nonparametric between-class scatter matrix S_b is defined as follows:

$$S_b = \sum_{i=1}^c P_i \sum_{j=1, j \neq i}^c \sum_{k=1}^{n_i} \frac{w_k^{(i,j)}}{n_i} \cdot (Z_k^{(i)} - M_j(Z_k^{(i)})) \cdot (Z_k^{(i)} - M_j(Z_k^{(i)}))^T, \quad (12)$$

where $Z_k^{(i)}$ is the k th sample in class i ; P_i is the prior probability of class i ; n_i is the number of samples in class i ; and $w_k^{(i,j)}$ is the distributed weight matrix of the k th sample in class i to the class j , which is defined as follows:

$$w_k^{(i,j)} = \frac{\text{dist}(Z_k^{(i)}, M_j(Z_k^{(i)}))^{-1}}{\sum_{t=1}^n \text{dist}(Z_i^{(i)}, M_j(Z_t^{(i)}))^{-1}}, \quad (13)$$

where $M_j(Z_k^{(i)})$ is the weighted mean of the sample $Z_k^{(i)}$. The weighted mean is defined as follows:

$$M_j(Z_k^{(i)}) = \sum_{t=1}^{n_i} \lambda_{kt}^{(i,j)} \cdot Z_t^{(j)}, \quad (14)$$

where $\lambda_{kt}^{(i,j)}$ is the weight of local mean. The weight of the local mean is defined as follows:

$$\lambda_{kt}^{(i,j)} = \frac{\text{dist}(Z_k^{(i)}, Z_t^{(j)})^{-1}}{\sum_{t=1}^{n_i} \text{dist}(Z_i^{(i)}, Z_t^{(j)})^{-1}}, \quad (15)$$

where $\text{dist}(a_1, a_2)$ is the Euclidean distance between vector a_1 and vector a_2 .

In equations (14) and (15), $Z_k^{(i)}$ has a weighted mean $M_j(Z_k^{(i)})$, the value of $M_j(Z_k^{(i)})$ is determined by each sample $Z_t^{(j)}$ in class j and $\lambda_{kt}^{(i,j)}$, and the value of $\lambda_{kt}^{(i,j)}$ is in reverse ratio to the Euclidean distance between $Z_k^{(i)}$ and $Z_t^{(j)}$. Therefore, the greater the distance between $Z_t^{(j)}$ and $Z_k^{(i)}$, the smaller the contribution of sample $Z_t^{(j)}$ for clustering.

2.4. Evaluation Indexes of the Clustering Effect. In this section, 3 kinds of clustering evaluation indexes are introduced: the partition coefficient (PC), classification entropy (CE), and clustering accuracy (Acc). They are defined as follows:

$$\begin{aligned} \text{PC} &= \left(\frac{1}{n}\right) \sum_{i=1}^c \sum_{k=1}^n u_{ik}^2, \\ \text{CE} &= -\left(\frac{1}{n}\right) \sum_{i=1}^c \sum_{k=1}^n u_{ik} \ln u_{ik}, \\ \text{Acc} &= \left(\frac{1}{n}\right) \sum_{i=1}^c \theta_i, \end{aligned} \quad (16)$$

where n is the number of the sample set and θ_i is the number of samples correctly partitioned into class i .

PC and Acc are closer to 1. CE is closer to 0, and so the clustering effect will be better.

3. Proposed Fault Diagnosis Method

In this section, a novel intelligent fault diagnosis method for bearings is discussed. The method includes 3 parts: DCAEN construction, improved ANW-GK clustering, and general procedure.

3.1. DCAEN Construction. The process of construction for the DCAEN is shown in Figure 2. The DCAEN is constructed by stacking CAE. The output of the pooling layer of the previous CAE serves as the input of the current CAE. At first, unlabeled data are used for pretraining the CAE layer by layer. Then, a full connection layer and softmax classifier are added to the coding part of the pretrained DCAEN, and a small number of labeled samples are used for the supervised fine-tuning of the network. Finally, the classification layer is removed from the fine-tuned network, and the trained DCAEN with better deep feature extraction capability is constructed.

In the process of pretraining the network, each layer of the DCAE becomes a shallow neural network, which can make use of the advantages of convex optimization of the shallow neural network and reduce the risk of the network falling into a local optimum. The pretrained network is fine-tuned by a few labeled data to achieve a better feature learning ability. Essentially, the process of encoding layer by layer is extracting abstract features step by step. With the increase of the layers, the features become more abstract and more global.

3.2. Improved ANW-GK Clustering. The membership degrees of the samples are used to calculate the corresponding cluster centers in the GK clustering algorithm. The different contributions of the samples, however, are not considered. Different samples should be given different feature weights, which can make the sample near the cluster center more typical. Thereby, the contribution of the typical sample which should play a leading role in the clustering process is increased. That is to say, when calculating the membership degree of sample Z_i belonging to class i , the samples near Z_i should belong to the same class and be given larger weights, while those farther away from Z_i should be given smaller weights.

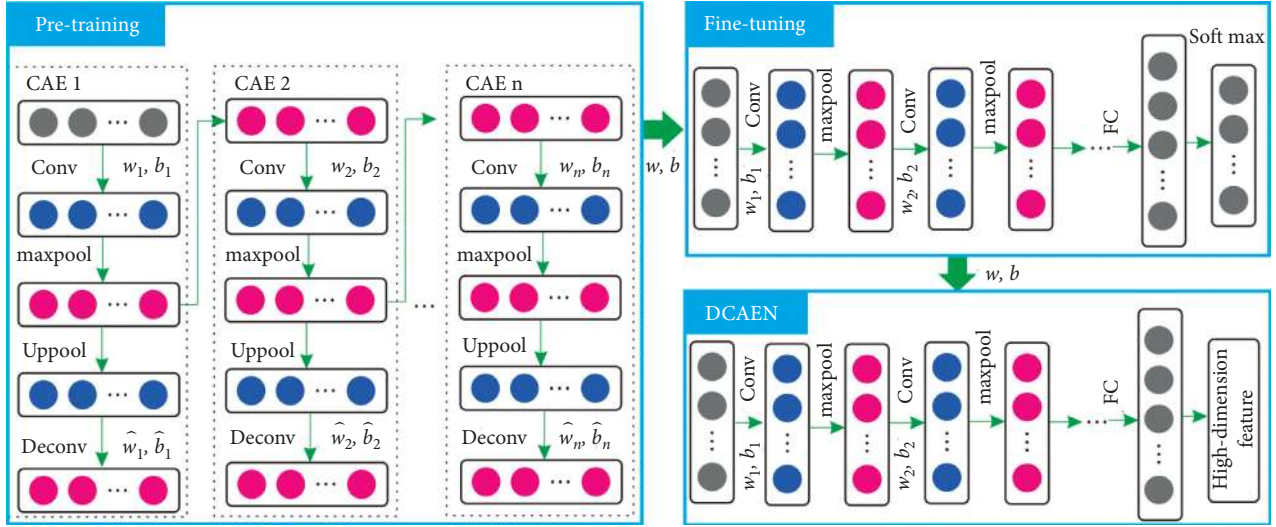


FIGURE 2: The structure of DCAEN.

Different weights of samples can be assigned in the NWFE. The importance of local information is emphasized. Therefore, in our method, the NWFE algorithm is integrated into GK clustering, and its new weighted clustering center \tilde{v}_i is defined as follows:

$$\tilde{v}_i = \frac{\sum_{k=1}^{n_i} \lambda_{ik} u_{ik}^m Z_k}{\sum_{t=1}^{n_i} \lambda_{it} u_{it}^m}. \quad (17)$$

When $Z_i = Z_k$, $\lambda_{ik} = 1$; when $Z_i \neq Z_k$, $\lambda_{ik} = \|Z_i - Z_k\|^{-1}$.

The objective function of NW-GK clustering is defined as follows:

$$J_{NW}(Z, \tilde{U}, \tilde{V}) = \sum_{i=1}^c \sum_{k=1}^{n_i} u_{ik}^m \cdot \tilde{D}_{ik}^2, \quad (18)$$

where $\tilde{D}_{ik}^2 = (Z_k - \tilde{v}_i)^T \cdot A_i \cdot (Z_k - \tilde{v}_i)$.

According to the Lagrange multiplier method, the weighted membership matrix can be updated as follows:

$$\tilde{u}_{ik} = \frac{1}{\sum_{g=1}^c \left(\tilde{D}_{ik} / \tilde{D}_{kg} \right)^{2/(m-1)}}, \quad (19)$$

until

$$\|\tilde{U}^{(l+1)} - \tilde{U}^{(l)}\| \leq \eta. \quad (20)$$

The cluster number c must be given in advance for the traditional GK clustering algorithm. It mainly depends on the experts' experience or relevant background knowledge. To enhance the adaptivity, the clustering evaluation PBMF function [28] is integrated into the NW-GK algorithm. According to the change of the PBMF function's value with the cluster number c , the optimal c can be selected. The PBMF function is defined as follows:

$$\text{PBMF}(c) = \left(\left(\frac{1}{c} \right) \cdot \left(\frac{E_1}{J_{NW}} \right) \cdot \max_{i,j=1}^c \|\tilde{v}_i - \tilde{v}_j\| \right)^2, \quad (21)$$

where E_1 is the value of J_{NW} when $c = 1$.

As can be seen from equation (21), the bigger the PBMF value, the better the clustering effect, and the value of the corresponding c is closer to the real number of clusters. Generally, $c \in [2, \sqrt{n}]$, where n is the number of samples.

Based on the above analysis, an ANW-GK clustering algorithm was developed, and its flowchart is shown in Figure 3.

- (1) The clustering parameters of c , c_{\max} , m , and η are initialized.
- (2) The fuzzy membership matrix U is initialized and satisfied $\mu_{ik} \in [0, 1]$, $\sum_{i=1}^c u_{ik} = 1$.
- (3) The weight matrix $\Lambda = \{\lambda_{ik}\}$ is calculated according to equation (15), and the cluster center \tilde{v}_i is updated according to equation (17).
- (4) The fuzzy membership matrix \tilde{U} is updated according to equation (19).
- (5) If $\Delta J_{NW} > \eta$, go to Step 2, and until the clustering information is converged. Otherwise, the next step is performed.
- (6) Calculate PBMF (c) according to equation (21), let $c = c + 1$, and go to Step 2. If $c > c_{\max}$, then a set of the PBMF values is obtained. Otherwise, return to Step 3.
- (7) The maximum value of PBMF (c) is found from the set of the PBMF values, and the corresponding c is the optimal cluster number. Its corresponding \tilde{U} and cluster centers \tilde{v} are the best clustering results.

3.3. Fault Diagnosis Procedure. According to the above-mentioned discussion, the proposed fault diagnosis method based on the DCAEN and ANW-GK is shown in Figure 4. The general procedure of the fault diagnosis method can be summarized as follows:

- (i) Step 1: the vibration signals are collected by a data acquisition system, and the collected signals are divided into training and test samples.

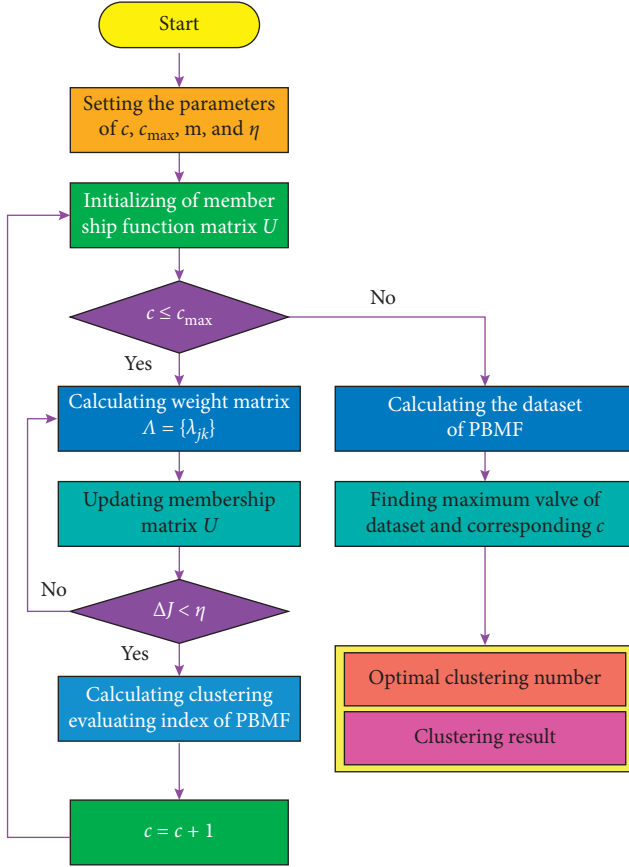


FIGURE 3: The flowchart of ANW-GK clustering.

- (ii) Step 2: the parameters of CAE1-CAEn are initialized, such as the convolution kernel size, pooling size, batch size, and learning rate.
- (iii) Step 3: CAE1-CAEn are unsupervised and pretrained by training samples, and the parameters are saved.
- (iv) Step 4: the encoders of the pretrained CAE1-CAEn are stacked, and a full-connection layer and a softmax classifier are added to the top. A small number of labeled training samples are used to fine-tune the pretrained network, and the parameters of the network are saved.
- (v) Step 5: the classifier is removed, and the trained DCAEN is constructed.
- (vi) Step 6: test samples are fed into the trained DCAEN, and the representative features are extracted. The main low-dimensional features of the representative features after the dimension reduction by t -SNE [31] are input into the ANW-GK algorithm for unsupervised fault recognition.
- (vii) Step 7: clustering accuracy, PC, and CE are employed to evaluate the clustering performance.

4. Experiment Verification and Analysis

Two cases of rolling bearing datasets are discussed in this section. They were used to validate the availability and superiority of the proposed fault diagnosis method.

4.1. Case 1: Bearing Dataset of Case Western Reserve University (CWRU)

4.1.1. Data Description. The data used for the verification of the proposed method were from the Case Western Reserve University (CWRU) bearing data center [32]. The data were collected by accelerometers from a motor driving mechanical system at a sampling frequency of 12 kHz. The motor bearings were seeded with faults using electrodischarge machining as shown in Figure 5. The system was able to bear 4 kinds of loads: 0–3 hp. Besides the normal (NR) operating status, single point fault with fault diameters of 0.007 in, 0.014 in, and 0.021 in were separately introduced at rolling element (BF), inner raceway (IF), and outer raceway (OF). Therefore, there were 10 categories of health conditions under a load in total. In this experiment, data with a load of 1 hp were used to make a sample set, each state included 100 samples, and each sample contained 1024 data points. For each fault category, 80 samples were randomly selected as training samples, and 20 samples were selected as test samples. The details of all the datasets are described in Table 1.

4.1.2. Parameters of the Model. We used the parameters of the DCAEN from several studies [18, 19]. The specific structural parameters were as follows: Conv₃₂⁶⁴ → max pool² → Conv₃₂³² → max pool² → Conv₁₆¹⁶ → max pool² → Conv₆₄³² → max pool² → Conv₆₄³² → max pool² → FC₂₀₀, where Conv_n^k denotes a convolutional layer with n convolution kernel of size $k \times 1$, and the default step size is 1; max pool² denotes a pooling layer with 2×1 size, and the default step size was 1. When pretraining, the minibatch was set to 80, the learning rate was set to 0.001, the epochs were set to 200, and the optimization algorithm was the Adam algorithm. When fine-tuning, the learning rate was set to 0.005 for improving the efficiency. The fuzzy weighted exponent m was set to 2, and the iteration termination tolerance η was set to 0.0001.

To determine the appropriate proportion of fine-tuned samples, 10%, 30%, 50%, and 70% of the training samples were used for fine-tuning, and each experiment was repeated 20 times. The statistical results are shown in Table 2. The performance indexes of each experiment are shown in Figure 6.

As can be seen from Table 2 and Figure 6, when the proportion of fine-tuned samples is 10%, the test accuracy of the model is up to 97.5%, the lowest is 88.5%; the maximum PC value is 0.875, and the minimum is 0.75; the minimum CE value is 0.32, and the maximum is 0.51. This indicates that the clustering evaluation index has large fluctuations, and the model stability is poor. As the proportion of fine-tuning samples increases, the stability of the model is increased, and the clustering indexes are improved gradually, but the improved magnitude is decreased gradually. So, under the premise of ensuring the performance of the model, the proportion of fine-tuning samples used in this paper was selected as 30%.

The cluster number was set to be $c \in [2, 14]$ ($c_{\max} = \sqrt{200}$), and the change of PBMF with c is shown in Figure 7. Therefore, the optimal number of clusters is $c = 10$, which encompasses the data of the actual situation.

TABLE 2: Statistics of test results under different proportion fine-tuning samples.

Fine-tuning sample ratio (%)	Evaluating index		
	Acc	PC	CE
10	0.94 ± 0.055	0.83 ± 0.05	0.41 ± 0.101
30	0.96 ± 0.015	0.88 ± 0.05	0.30 ± 0.10
50	0.97 ± 0.01	0.90 ± 0.012	0.26 ± 0.02
70	0.98 ± 0.01	0.90 ± 0.023	0.25 ± 0.031

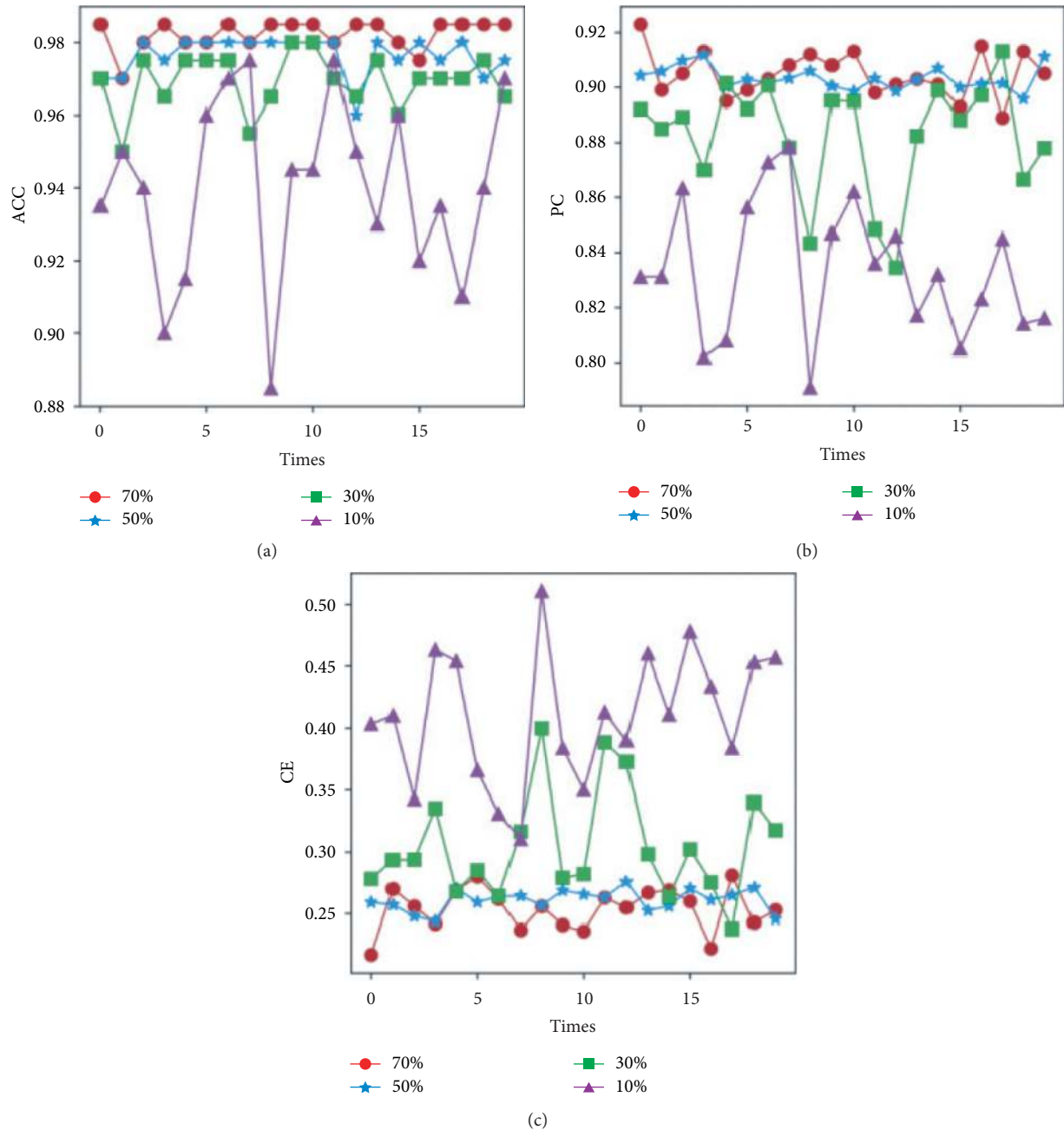


FIGURE 6: Test results with different proportions of fine-tuned samples.

4.1.3. Results and Analysis. The established DCAEN was first pretrained layer by layer, and the pretrained DCAEN was fine-tuned by 30% of the labeled training samples. Test samples were input into the fine-tuned

DCAEN, and the 200×200 high-dimensional feature set was obtained. For visualization, the t-SNE was used to reduce the 200×200 feature set to a 200×2 feature set, which was used as the input of the ANW-GK clustering

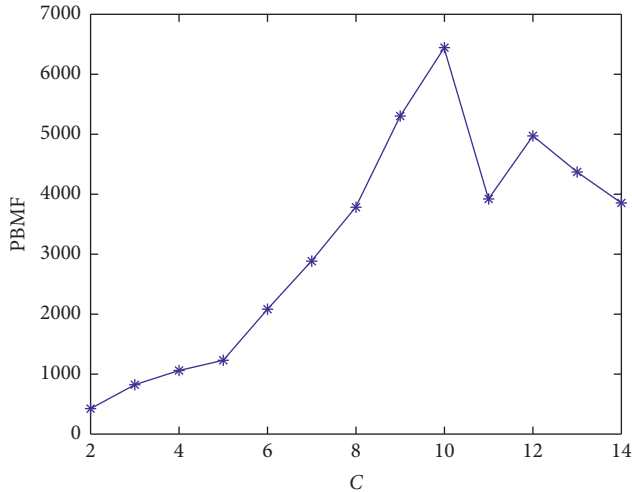


FIGURE 7: The change of PBMF with cluster number c .

algorithm. The results of clustering are shown in Figure 8.

In Figure 8, V1 to V10 correspond to the cluster centers of 10 bearing states, and their specific coordinate values are shown in Table 3. It can be seen from Figure 8 that 10 bearing states are clearly separated and gathered in the vicinity of the cluster centers. Each group of samples is packed tight, and the spaces between the classes are large, and no aliasing occurs. The average memberships of each group of samples are shown in Table 4. It can be seen that the average membership of the NR group samples for V3 is 0.982, which is much larger than the other 9 cluster centers. Therefore, the NR group samples belong to the V3 class. Similarly, the NR-OF3 samples belong to V9, V1, V4, V6, V7, V5, V8, V10, and V2, respectively. Therefore, the proposed method has obvious fault identification effects. It should be noted that the membership degree of the BF3 group belonging to the V4 class is 0.735, and memberships belonging to V5, V9, and V7 are 0.094, 0.074, and 0.047, respectively, which are significantly higher than the membership degrees of other cluster centers. Therefore, the BF3 group samples are mainly affected by the IF1, BF1, and IF2 samples when clustering. Similarly, it can be seen from the memberships of the OF3 group samples that this group of samples is greatly affected by the OF1 group. This is consistent with the conclusion of Figure 8.

4.1.4. Generalization Performance under Different Loads.

In practical applications of mechanical equipment, the loads of bearings are often variable. In this section, we discuss the generalization performance of the proposed method under different loads. The model was trained and fine-tuned using the training set of 1 hp load. The test sets were under the loads of 0 hp, 2 hp, and 3 hp. The experimental results are shown in Figure 9. Under 3 different loads, the clustering accuracies are 96%, 97%, and 95.5%; the PC values are 0.811, 0.853, and 0.879, respectively; and the CE values are 0.473, 0.399, and 0.312, respectively. The clustering results still maintained a high precision. For the stability of the model,

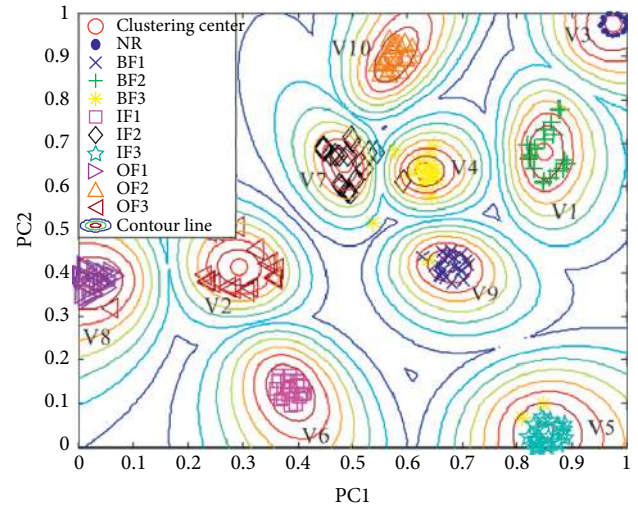


FIGURE 8: The cluster contour map of the proposed method.

each experiment was repeated 20 times, and the statistical results are shown in Figure 10. Under different loads, the clustering accuracy rate is above 96%, the PC value is above 0.8, and the CE value is within 0.5. These results show that the proposed method has a certain generalization when the load changes.

4.1.5. Comparative Experiment. To illustrate the superiority of the proposed method, the following comparative experiments were conducted. (1) We compared our proposed method with traditional signal processing and handcraft feature methods. The original vibration signals were decomposed into several IMF components using the EMD, and permutation entropy (PE, $m=2$, $r=0.1$ SD) was employed to calculate the entropy value of each IMF component as feature vectors. For visualization, t-SNE was used for the dimension reduction. The 2-dimensional IMF-PE vectors were input into the ANW-GK cluster for fault identification. The multifractal method was used to extract the features of the original vibration signals, and the q -D (q) parameters ($q=10$) of the signal were used as the feature vectors. For visualization, t-SNE was used for the dimension reduction. The 2-dimensional q -D (q) feature vectors were input into the ANW-GK cluster for fault identification. (2) We compared our proposed method with the SDAE. The SDAE was used to extract the features from the original vibration signal, and the input ANW-GK clustering was for fault identification. To maintain consistency, the network structure of the SDAE was 1024-1024-96-192-192-200. (3) We compared our proposed method with the GK clustering algorithm. The DCAEN was used to extract the features of the original vibration signal, and the extracted features were input into the GK clustering algorithm for fault identification. The comparison results are shown in Table 5, Figure 8, and Figure 11.

In comparison to the manual feature extraction methods of the EMD + FE and multifractal methods, features learned by the DCAEN have a better cluster recognition effect, as

TABLE 3: Cluster centers of 10 types of signals.

Sample group		NR	BF1	BF2	BF3	IF1	IF2	IF3	OF1	OF2	OF3
Cluster	x	0.976	0.674	0.85	0.63	0.385	0.485	0.851	0.289	0.576	0.294
Centers	y	0.975	0.419	0.679	0.635	0.132	0.65	0.034	0.38	0.895	0.413

TABLE 4: Membership degrees of 10 types of signals.

Sample group	Membership degree									
	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
NR	0.008	0.001	0.982	0.003	0.001	0.000	0.001	0.001	0.001	0.002
BF1	0.010	0.006	0.003	0.017	0.005	0.004	0.012	0.002	0.938	0.003
BF2	0.884	0.005	0.030	0.036	0.005	0.003	0.009	0.003	0.017	0.008
BF3	0.012	0.009	0.005	0.735	0.094	0.005	0.047	0.003	0.074	0.016
IF1	0.001	0.008	0.001	0.002	0.003	0.973	0.003	0.003	0.004	0.002
IF2	0.007	0.015	0.005	0.121	0.003	0.006	0.761	0.006	0.023	0.053
IF3	0.004	0.002	0.001	0.002	0.978	0.003	0.002	0.001	0.006	0.001
OF1	0.001	0.010	0.001	0.002	0.001	0.004	0.002	0.975	0.002	0.002
OF2	0.004	0.002	0.004	0.011	0.001	0.001	0.012	0.002	0.003	0.960
OF3	0.006	0.724	0.005	0.023	0.007	0.057	0.032	0.097	0.025	0.024

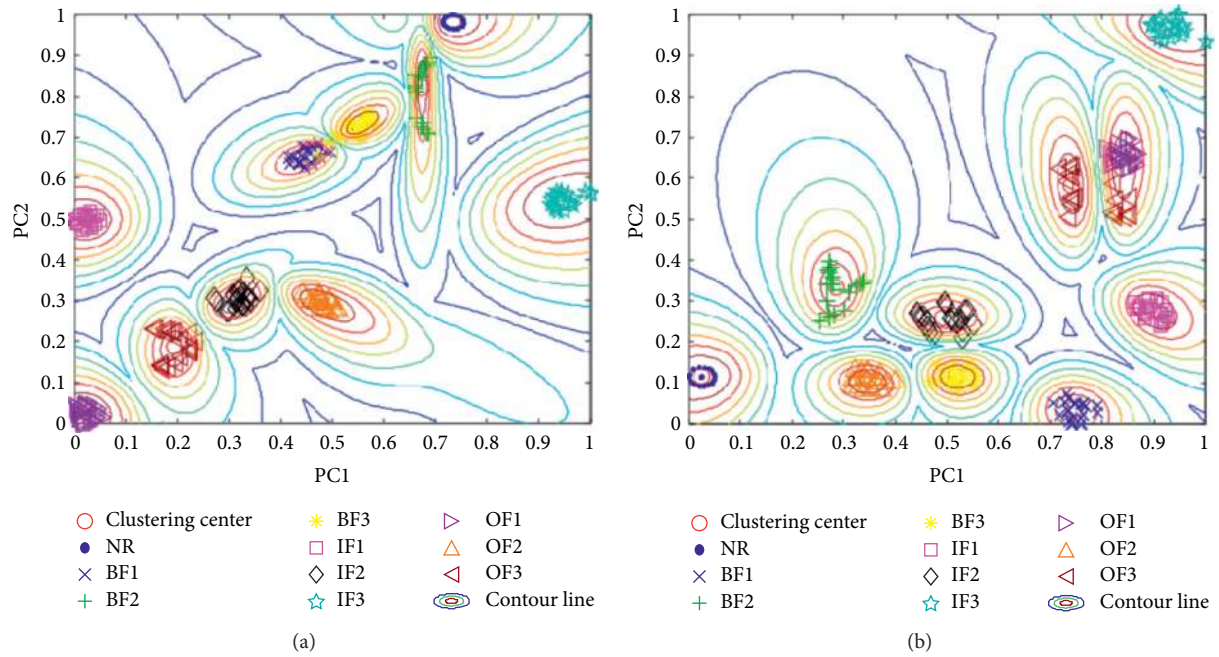


FIGURE 9: Continued.

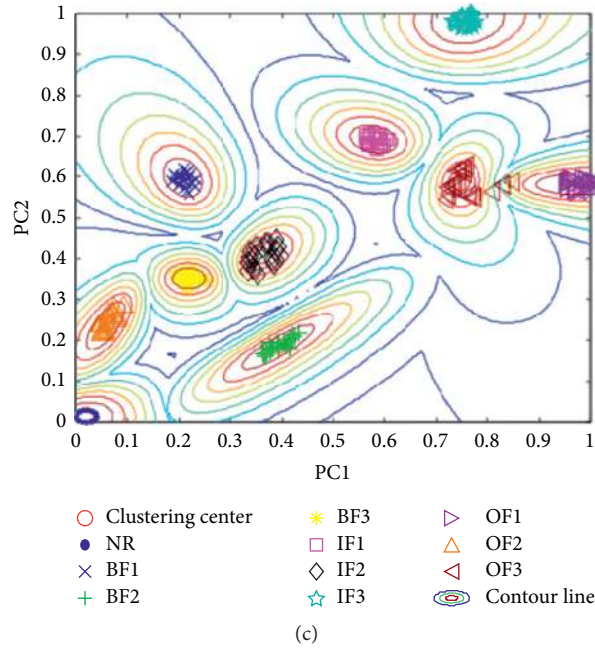


FIGURE 9: Recognition results under different loads. (a) 0 hp; (b) 2 hp; (c) 3 hp.

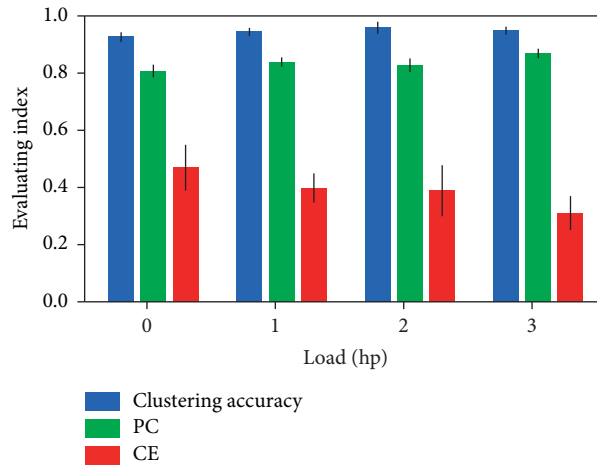


FIGURE 10: The performances of the proposed method under different loads.

TABLE 5: The recognition results of different methods.

Methods	Accuracy (%)									Evaluating index			
	NR	BF1	BF2	BF3	IF1	IF2	IF3	OF1	OF2	OF3	AVG	PC	CE
EMD + PE + ANW-GK	100	100	15	60	100	80	0	90	90	100	73.5	0.733	0.535
Multifractal + ANW-GK	100	70	90	100	100	50	10	100	90	100	81	0.760	0.526
SDAE + ANW-GK	100	100	65	80	95	100	100	100	85	55	88	0.775	0.54
DCAEN + GK	100	100	100	80	100	95	100	100	100	80	95.5	0.836	0.413
Proposed method	100	100	100	80	100	95	100	100	100	90	96.5	0.848	0.399

shown in Table 5. For features extracted using the EMD + PE, “OF1” and “OF3” are seriously aliased and “BF2” and “BF3” are seriously aliased, as shown in Figure 11(a); for features extracted using the multifractal method, “IF3” and “OF3” are seriously aliased and “BF1” and “IF2” are seriously aliased, as shown in Figure 11(b). This is mainly because the

features extracted manually are not comprehensive, and important sensitive features may be lost, which results in identification difficulties.

As shown in Table 5, compared with SDAE, the cluster recognition effect of the features learned using the DCAEN is also better. For features learned by SDAE, “BF1” and

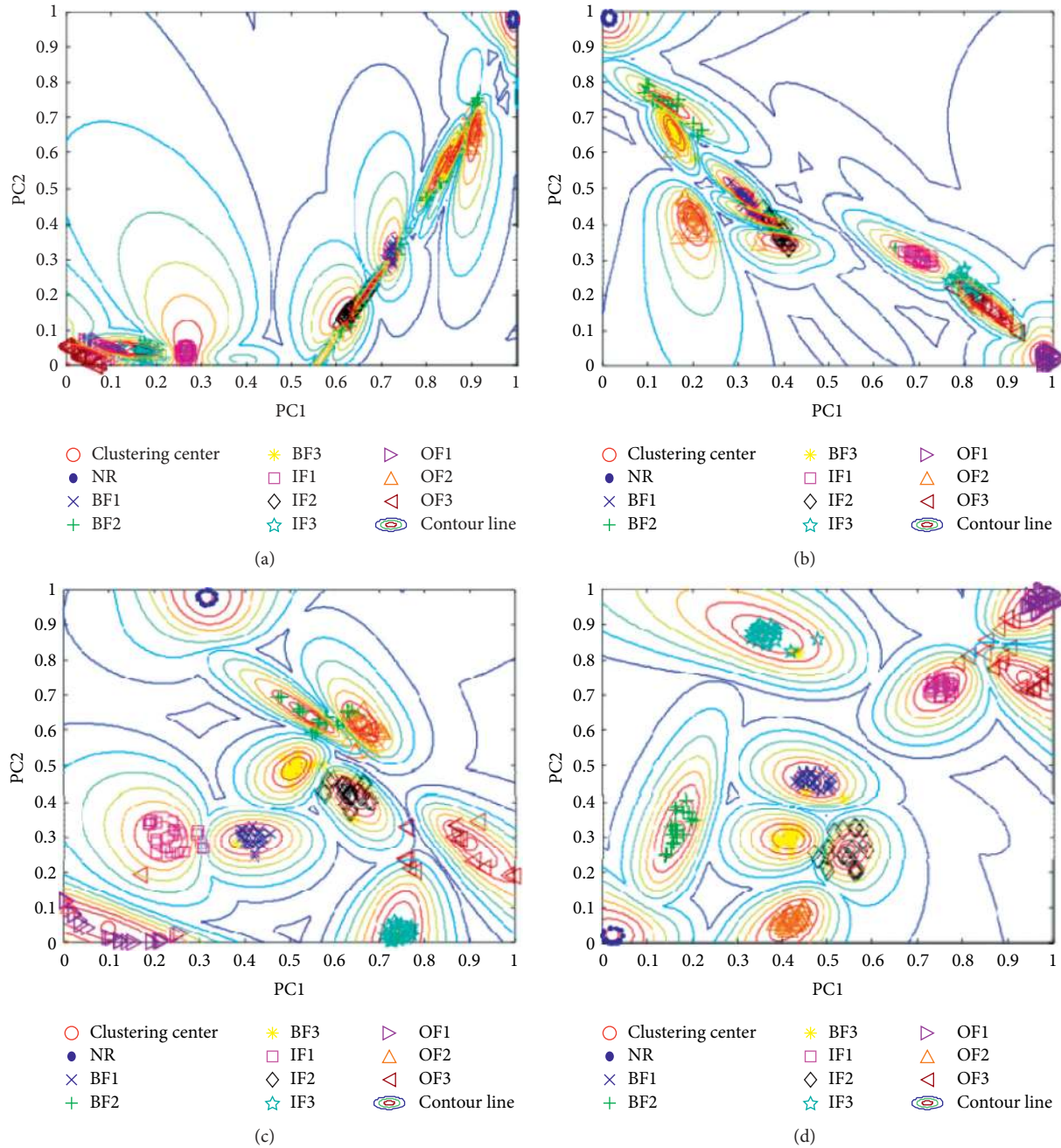


FIGURE 11: The results of the contrast tests. (a) EMD+FE+AWN-GK; (b) Multifractal+ANW-GK; (c) SDAE+AWN-GK; (d) DCAEN+GK.

“OF2” are aliased, and “OF3” and “IF3” are aliased, as shown in Figure 11(c). The full connection between each network layer was used in the SDAE, which results in a large amount of redundancy in the network’s structural parameters. This makes the features learned by the network more global, while the locality of the features may be ignored. The structure of the convolution and pooling plus full connection layer is used in the DCAEN. The convolution pooling layer learns the local features from the input data, and the full connection layer learns the global features. Thus, the features extracted by our method are more distinguishable.

Compared with GK clustering, the fault recognition effect of the ANW-GK clustering is much better, as shown in Table 5. When GK clustering was used to identify the fault types of features learned by the DCAEN, “OF1,” “IF1,” and “OF3” are slightly aliased, as shown in Figure 11(d); when our method is used to identify the fault types of features learned by the DCAEN, no aliasing between the various types occurs, as shown in Figure 8. This is mainly because different weights are given to each sample in our method, which enhances the role of typical samples. The different importance of samples for each type is more

effectively characterized, so that the clustering accuracy is improved.

4.2. Case 2: Laboratory-Simulated Bearing Fault Dataset. To further verify the effectiveness of the proposed method, the proposed method was applied to analyze the laboratory-simulated bearing faults dataset.

4.2.1. Experimental Setup. The laboratory-simulated bearing fault data were collected from a rotor test bench and shown in Figure 12. The rotor test bench was used to simulate different operating states of the ball bearings. A three-phase inverter motor, a shaft, and a speed controller were used to vary speeds of the test bearings. The single point fault is arranged on the bearings (NSK6308) using a wire electrical discharge machine and a file as shown in Figure 13. A couple of accelerometers (HD-YD232) were placed vertically on the bearing seat to collect the vibration signals of the test bearings.

The dataset included five different operating states of the bearings: normal (NR), outer ring fault (OF), inner ring fault (IF), rolling element fault (BF), and fix fault (FF). In the experiment, the rotating speed of the shaft was 2600 rpm, the sampling frequency was 8 kHz, and 200 samples were collected in each operating state. For each operating condition, 160 samples were randomly selected as training samples, and 40 samples were selected as test samples. The samples used for training a deep network must contain at least one complete signal period; otherwise, fault features cannot be effectively learned. To meet this requirement, the sample length must be longer than the number of points contained in a complete period, and the latter can be calculated by the sampling frequency and bearing speed. Since the number of data points in the collected raw data is constant, the sample length is inversely proportional to the number. If the sample length is too long, on the one hand, the number of samples may be too small and affect the training of the model; on the other hand, it may increase the cost of computation and affect the training speed. On the premise of ensuring the training effect, for the convenience of storage and calculation, the length of the sample is 2048. The details of all the datasets are described in Table 6.

4.2.2. Parameters of the Model. The parameters of the model in Case 1 were used. The cluster number was set to $c \in [2, 14]$ ($c_{\max} = \sqrt{200}$), and the change of PBMF with c is shown in Figure 14. Therefore, the optimal cluster number is $c = 5$.

4.2.3. Results and Analysis. The training samples were input into the established DCAEN for unsupervised pretraining, and 30% of the labeled training samples were used for fine-tuning. We input 200 ($5 \times 40 = 200$) test samples into the trained DCAEN and obtained the 200×200 dimensional features. For visualization, t-SNE was used to reduce the 200-dimensional features to 2 dimensions. The two-dimensional features were input into ANW-GK clustering for fault identification. The

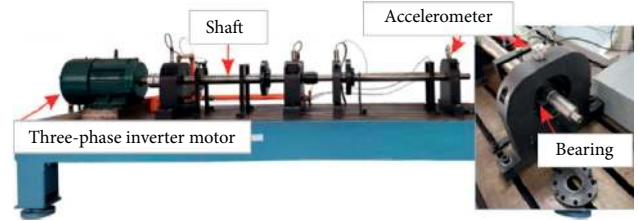


FIGURE 12: The test bench of the bearings.

clustering accuracy, PC value, and CE value are 97.5%, 0.915, and 0.186, respectively. The results of clustering are shown in Figure 15.

In Figure 15, V1, V2, V3, V4, and V5 are the cluster centers of FF, NR, BF, IF, and OF, respectively, and their specific coordinate values are shown in Table 7. As can be seen from Table 7 and Figure 15, five kinds of samples are clearly separated and clustered near their cluster centers. Different types of samples are gathered closely, no aliasing occurs and the distances between classes are large. The average membership degrees of each group of samples are shown in Table 8. The membership of the first group of samples for V2 is much larger than that of the other four groups, which indicates that the first group of samples belongs to V2. Similarly, the other groups of samples belong to different classes. Therefore, the excellent fault identification effect of the proposed method is verified again.

4.2.4. Generalization Performance under the Different Rotating Speeds. In actual mechanical equipment, the rotating speeds of bearings are often variable. Consequently, the generalization performances of the proposed method at different rotating speeds were tested. The model was trained with the training set of 2600 rpm. The test sets were at rotating speeds of 2800 rpm, 3000 rpm, and 3200 rpm. The fault diagnosis results are shown in Figure 16. At three different speeds, the clustering accuracies are 97%, 96.5%, and 98.5%; the PC values are 0.912, 0.901, and 0.922, respectively; the CE values are 0.194, 0.215, and 0.14, respectively. It can be seen that the proposed method still maintains higher fault diagnosis accuracy at variable rotating speeds. Therefore, our method has certain generalization performances at different rotating speeds.

To avoid a contingency, the experiments at different rotating speeds were performed 20 times. The final results are the average of clustering evaluation indexes for the 20 experiments, as shown in Figure 17. At different rotating speeds, the clustering accuracy was maintained above 96%, the PC value was above 0.85, and the CE value was below 0.25. These results show that the proposed method has a good fault identification ability at different rotating speeds.

5. Results and Discussion

As mentioned above, we know that it is difficult and sometimes impossible to obtain a large number of labeled samples in the process of fault diagnosis. The insufficiency of labeled samples easily leads to lower diagnostic accuracy.

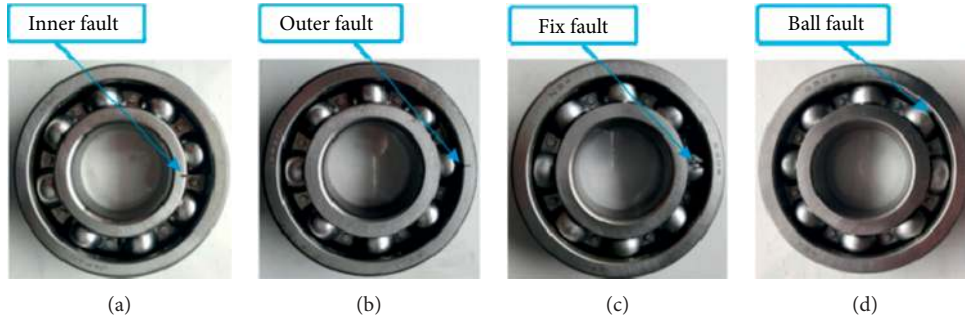


FIGURE 13: The faults of bearing in four locations: (a) inner fault; (b) outer fault; (c) fix fault; (d) ball fault.

TABLE 6: Description of the measured datasets.

Fault type		NR	Ball	Cage	Inner	Outer	Speed
Category label		1	2	3	4	5	
Fault diameter (mm)		0	0.5	3	0.2	0.2	2600
Dataset	Train	160	160	160	160	160	2800
Size	Test	40	40	40	40	40	3000
							3200

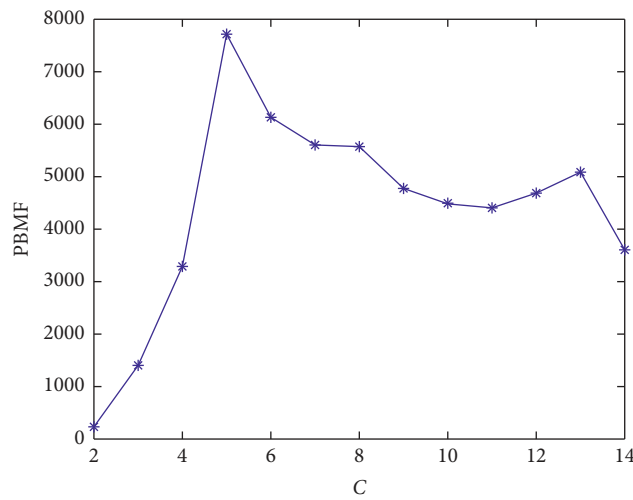


FIGURE 14: The change of PBMF with cluster number c .

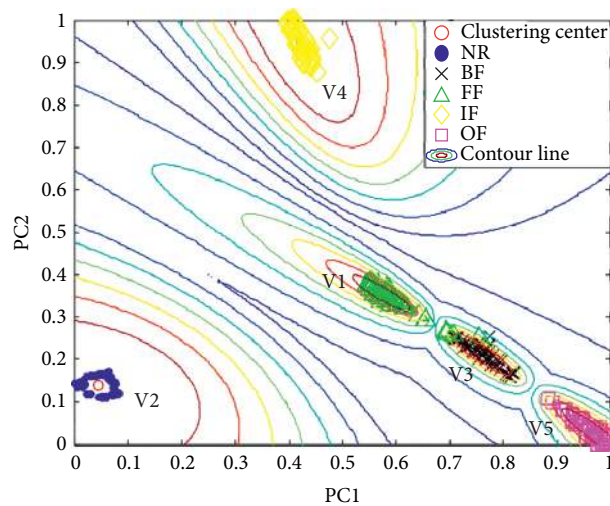


FIGURE 15: The cluster contour map of the proposed method.

TABLE 7: The cluster centers for five types of signals.

Sample group		NR	BF	FF	IF	OF
Cluster	x	0.046	0.756	0.585	0.423	0.953
Center	y	0.139	0.216	0.348	0.946	0.043

TABLE 8: The membership degree of five types of signals.

Sample group	Membership degree				
	V1	V2	V3	V4	V5
NR	0.001	0.997	0.001	0.000	0.001
BF	0.066	0.002	0.868	0.003	0.061
FF	0.855	0.002	0.123	0.004	0.016
IF	0.001	0.000	0.001	0.996	0.002
OF	0.012	0.001	0.045	0.002	0.94

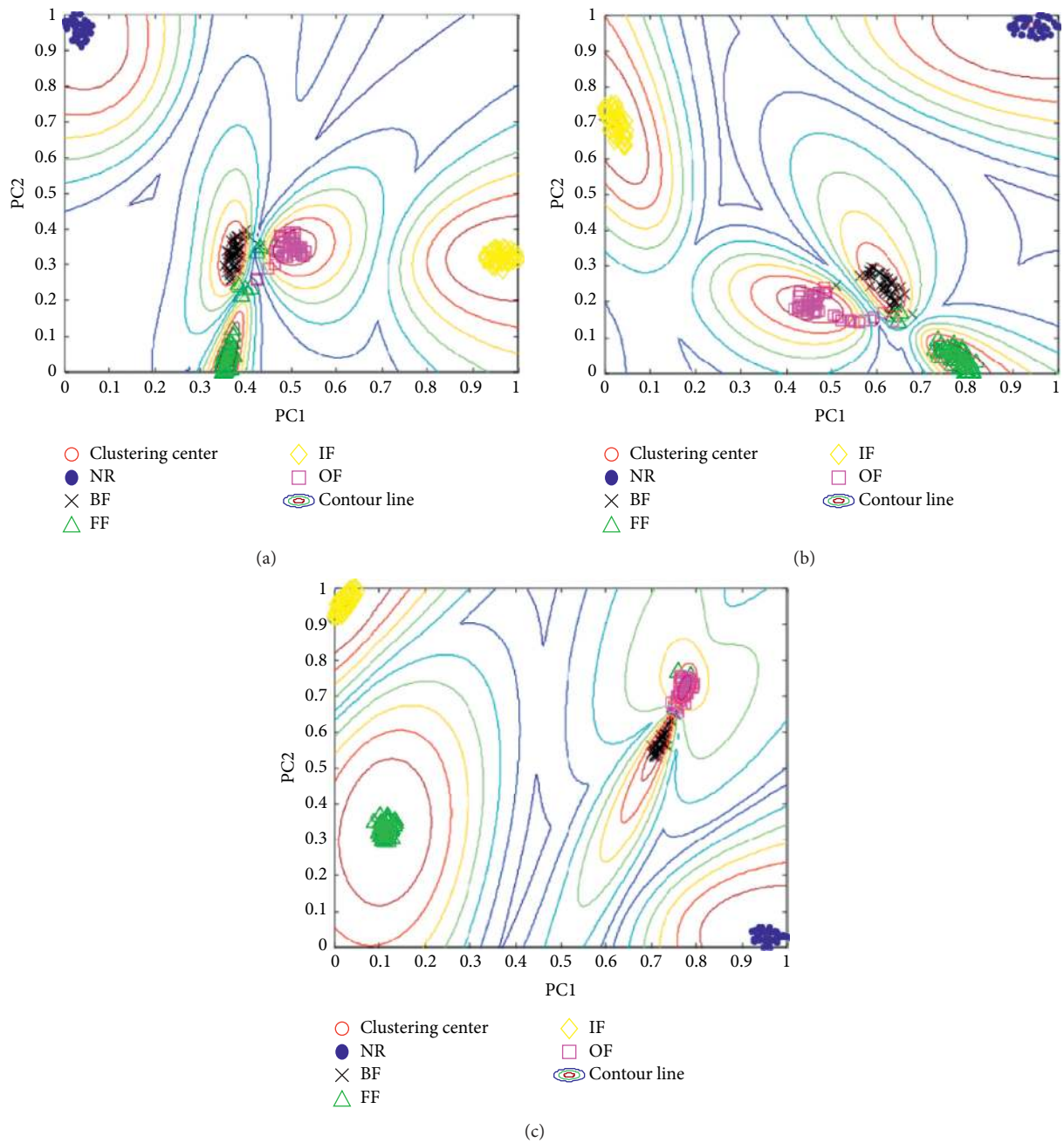


FIGURE 16: Recognition results at different rotating speeds. (a) 2800 rpm; (b) 3000 rpm; (c) 3200 rpm.

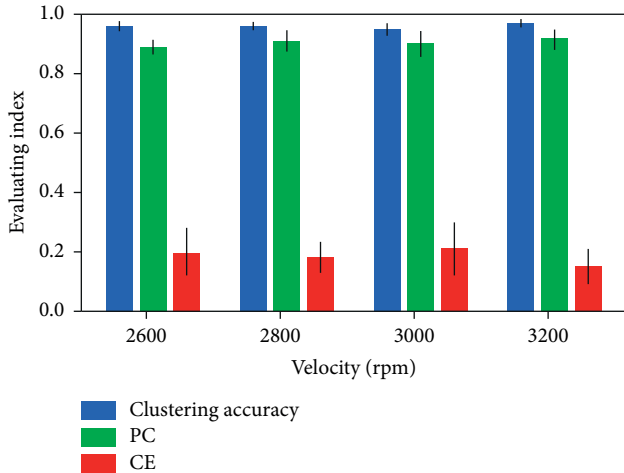


FIGURE 17: The performances of the proposed method at different speeds.

Therefore, it is important to explore fault diagnosis methods that use fewer labeled samples to achieve higher accuracy. In this paper, an intelligent fault diagnosis method using DCAEN and ANW-GK clustering is proposed. The method can identify fault types with a few labeled samples. The performance of the method is validated on two bearing datasets. However, there are still some potential problems and research directions remained to be improved and studied.

To effectively use a small number of labeled samples to improve the feature extraction capability of the model, a labeled sample fine-tuning technique is used during the construction of DCAEN. Through experiments, 30% of the training samples are used to fine-tune DCAEN and the model can obtain better diagnostic performance. But, this is only tested on two datasets with single fault, which has certain limitations.

The parameters optimization of DCAEN needs to be considered. The number of convolutional layers, the size of the convolution kernel, the pooling size, and the activation function have an important impact on the performance of model. Based on the empirical values in the references, there are many shortcomings for model performance. Therefore, the issue of how to choose parameters of DCAEN should be considered.

ANW-GK clustering has certain advantages compared to the existing method (i.e., GK clustering). The integration of NWFE and PBMF into GK clustering improves the algorithm's fault identification ability and also increases its complexity. Whether it affects the real-time performance of fault diagnosis method is required be studied lately.

6. Conclusions

A method based on the DCAEN and ANW-GK clustering for rolling bearing fault diagnosis is proposed in this paper. In our method, the fine-tuned DCAEN with a few labeled samples was used to extract high-level features of the input signals, and the extracted features reduced in dimension by t-SNE were input into the improved ANW-GK clustering

algorithm for fault identification. Our method was validated on a benchmark bearing dataset and a laboratory-measured bearing dataset. The diagnostic accuracies are 96.5% and 97.5%, the PC values are 0.848 and 0.915, and the CE values are 0.399 and 0.186. The experimental results show that the feature extraction is better than that of other models, such as the EMD + PE/multifractal/SDAE model. The classification accuracies also show that the ANW-GK clustering can identify the bearing faults effectively under various conditions.

In the future, we will focus on deep embedded clustering that directly adds a clustering layer to the top of the DCAEN. The deep embedded clustering will iteratively improve the weight parameters and clustering goals of the joint optimization network through soft allocation. Thus, the model will not need to be fine-tuned, and the operation efficiency can also be improved.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no potential conflicts of interest with respect to the publication of this article.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (51675253) and the National Key Research and Development Program of China (2016YFF0203303-04).

References

- [1] Y. Miao, M. Zhao, J. Lin, and Y. Lei, "Application of an improved maximum correlated kurtosis deconvolution method for fault diagnosis of rolling element bearings," *Mechanical Systems and Signal Processing*, vol. 92, pp. 173–195, 2017.
- [2] H. Shao, H. Jiang, H. Zhang, W. Duan, T. Liang, and S. Wu, "Rolling bearing fault feature learning using improved convolutional deep belief network with compressed sensing," *Mechanical Systems and Signal Processing*, vol. 100, pp. 743–765, 2018.
- [3] A. Diez-Olivan, J. Del Ser, D. Galar, and B. Sierra, "Data fusion and machine learning for industrial prognosis: trends and perspectives towards industry 4.0," *Information Fusion*, vol. 50, pp. 92–111, 2018.
- [4] M. Alabsi, Y. Liao, and A. A. Nabulsi, "Bearing fault diagnosis using deep learning techniques coupled with handcrafted feature extraction: a comparative study," *Journal of Vibration and Control*, Article ID 1077546320929141, 2020.
- [5] Q. Q. Chen, S. W. Dai, and H. D. Dai, "A rolling bearing fault diagnosis method based on EMD and quantile permutation entropy," *Mathematical Problems in Engineering*, vol. 2019, Article ID 3089417, 8 pages, 2019.
- [6] H. Li, X. T. Li, R. Jia, L. Xingqi, and Z. Jixing, "Fault diagnosis for hydropower units based on fractal and probabilistic neural

- network,” *Journal of Hydroelectric Engineering*, vol. 38, no. 3, pp. 92–100, 2019.
- [7] F. Jia, Y. Lei, J. Lin, X. Zhou, and N. Lu, “Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data,” *Mechanical Systems and Signal Processing*, vol. 72-73, pp. 303–315, 2016.
- [8] R. Liu, B. Yang, E. Zio, and X. Chen, “Artificial intelligence for fault diagnosis of rotating machinery: a review,” *Mechanical Systems and Signal Processing*, vol. 108, pp. 33–47, 2018.
- [9] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, “Deep learning and its applications to machine health monitoring,” *Mechanical Systems and Signal Processing*, vol. 115, pp. 213–237, 2019.
- [10] S. Khan and T. Yairi, “A review on the application of deep learning in system health management,” *Mechanical Systems and Signal Processing*, vol. 107, pp. 241–265, 2018.
- [11] Y. Lecun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [12] J. Schmidhuber, “Deep learning in neural networks: an overview,” *Neural Networks*, vol. 61, pp. 85–117, 2015.
- [13] Z. Q. Chen, C. Li, and R. V. Sanchez, “Gearbox fault identification and classification with convolutional neural networks,” *Shock and Vibration*, vol. 2015, Article ID 390134, 10 pages, 2015.
- [14] G. Jiang, H. He, P. Xie, and Y. Tang, “Stacked multilevel-denoising autoencoders: a new representation learning approach for wind turbine gearbox fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, vol. 66, no. 9, pp. 2391–2402, 2017.
- [15] C. Lu, Z.-Y. Wang, W.-L. Qin, and J. Ma, “Fault diagnosis of rotary machinery components using a stacked denoising autoencoder-based health state identification,” *Signal Processing*, vol. 130, pp. 377–388, 2017.
- [16] Z. Chen and Z. Li, “Fault diagnosis method of rotating machinery based on stacked denoising autoencoder,” *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 6, pp. 3443–3449, 2018.
- [17] O. Janssens, V. Slavkovicj, B. Vervisch et al., “Convolutional neural network based fault detection for rotating machinery,” *Journal of Sound and Vibration*, vol. 377, pp. 331–345, 2016.
- [18] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, “A new deep learning model for fault diagnosis with good anti-noise and domain adaptation ability on raw vibration signals,” *Sensors*, vol. 17, no. 2, pp. 425–446, 2017.
- [19] L. Jing, M. Zhao, P. Li, and X. Xu, “A convolutional neural network based feature learning and fault diagnosis method for the condition monitoring of gearbox,” *Measurement*, vol. 111, pp. 1–10, 2017.
- [20] G. Jiang, H. He, J. Yan, and P. Xie, “Multiscale convolutional neural networks for fault diagnosis of wind turbine gearbox,” *IEEE Transactions on Industrial Electronics*, vol. 66, pp. 3196–3207, 2018.
- [21] C. Li, J. V. De Oliveira, M. Cerrada, D. Cabrera, R. V. Sanchez, and G. Zurita, “A systematic review of fuzzy formalisms for bearing fault diagnosis,” *IEEE Transactions on Fuzzy Systems*, vol. 27, no. 7, pp. 1362–1382, 2019.
- [22] A. Davari, M. Hamiruce Marhaban, S. B. M. Noor, and A. Karimoddini, “Parameter estimation of K-distributed sea clutter based on fuzzy inference and Gustafson-Kessel clustering,” *Fuzzy Sets and Systems*, vol. 163, no. 1, pp. 45–53, 2011.
- [23] A. Benyounes, A. Hafaiifa, and M. Guemana, “Gas turbine modeling based on fuzzy clustering algorithm using experimental data,” *Applied Artificial Intelligence*, vol. 30, no. 1, pp. 29–51, 2016.
- [24] Y. Hua, H. Jiang, H. Tian, X. Xu, and L. Chen, “A comparative study of clustering analysis method for driver’s steering intention classification and identification under different typical conditions,” *Applied Sciences*, vol. 7, no. 10, p. 1014, 2017.
- [25] C. Li, M. Cerrada, D. Cabrera et al., “A comparison of fuzzy clustering algorithms for bearing fault diagnosis,” *Journal of Intelligent & Fuzzy Systems*, vol. 34, no. 6, pp. 3565–3580, 2018.
- [26] D. Chen, Y. D. Zhang, C. Y. Yao et al., “Fault diagnosis based on FVMMD multi-scale permutation entropy and GK fuzzy clustering,” *Journal of Mechanical Engineering*, vol. 54, no. 1, pp. 16–27, 2018.
- [27] B. C. Kuo and D. A. Landgrebe, “Nonparametric weighted feature extraction for classification,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 5, pp. 1096–1105, 2004.
- [28] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik, “Validity index for crisp and fuzzy clusters,” *Pattern Recognition*, vol. 37, no. 3, pp. 487–501, 2004.
- [29] X. L. Zhang, X. Zhou, and C. Tang, “A deep convolutional auto-encoding neural network and its application in bearing fault diagnosis,” *Journal of Xi’an Jiaotong University*, vol. 52, no. 7, pp. 1–8, 2018.
- [30] F. Qian, M. Yin, X. Y. Liu et al., “Unsupervised seismic facies analysis via deep convolutional autoencoders,” *Geophysics*, vol. 83, no. 3, pp. 39–43, 2018.
- [31] L. V. D. Maaten and G. Hinton, “Visualizing data using t-SNE,” *Journal of Machine Learning Research*, vol. 9, no. 2605, pp. 2579–2605, 2008.
- [32] Case Western Reserve University, “Bearing data center website,” 2018, <https://csegroups.case.edu/bearingdatacenter/home>.